

Impact of the face registration techniques on facial expressions recognition

B.Allaert¹, J.Mennesson¹, IM.Bilasco¹, C.Djeraba¹

^aUniv. Lille, CNRS, Centrale Lille, UMR 9189 - CRISTAL -
Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France

Abstract

Recent methodologies for facial expression recognition have been proposed and have obtained good results in near-frontal view. However, these situations do not fairly represent in-the-wild challenges, where expressions are natural and the subject is free of its movement. This is reflected in the accuracy drop of facial expression methods obtained on recent databases. Two challenges (head pose variations and large displacements) in facial expression recognition are studied in this paper. Experiments are proposed in order to quantify the impact of free head movements using representative expression recognition approaches (LBP, LBP-TOP, HOOF). We propose an experimental protocol (SNaP-2DFe) that records, under controlled light, facial expressions with two cameras: one attached on the head and one placed in front of the subject. As in both cameras facial expressions are the same, differences in performances measured on each camera show the impact of head pose variations and large displacements on the underlying recognition approach.

Keywords: Facial expressions, Head pose variations, Large displacements

1. Introduction

Facial expression recognition has attracted great interest over the past decade in various domains. Given the significant role of the face in human communication, several researches have been conducted on facial expression recognition in various contexts.

Several systems evaluate their performances on image collections, where facial expressions are played by actors, in order to obtain exaggerated facial deformations (acted expressions). Several approaches [?] obtain very good results in these settings. However, these collections do not fairly represent in-the-wild challenges, where expressions are natural (spontaneous expressions), and problems like head pose variations and large displacements are frequent, as illustrated in Figure ?. To answer these challenges, recently created collections [?] are mainly related to interaction situations where people are free of their movements. They are more challenging due to misalignment in faces, primarily caused by head motions, but also, spontaneous expressions.

State-of-the-art approaches that provide good results in near-frontal view have evolved in order to improve their robustness in the presence of head motions. The most commonly used

*Corresponding author



Figure 1: Faces captured in-the-wild, from GENKI-4K database [?].

solution to deal with head motions is to add a pre-processing step generally based on face registration in order to obtain frontal faces [? ?]. However, these methods casually induce texture changes that are not related to the underlying expression.

As in-the-wild settings, expressions are not acted, their intensity is getting smaller, and, hence, the changes induced by the registration interfere with changes induced by the expression itself. Indeed, spontaneous facial expressions are quite different from acted expressions in terms of facial movement amplitudes and/or texture changes. This makes them more difficult to characterize. In this context, systems based on dynamic textures may provide better performance [? ?]. Indeed, they detect subtle changes occurring on the face and do not require large changes in appearance, as texture-based or geometry-based approaches expect. However, these approaches are much more sensitive to varying head motion.

The question about the use and the impact of registration approaches arises especially when facial expression analysis is done in uncontrolled context. The use of registration approaches is increasing, despite a lack of evidence about their effectiveness due to the heterogeneity of the databases.

In this study, we address two challenges : head pose variations and large displacements in facial expressions recognition, denoted HPV and LD, respectively. In section ??, we discuss the impact of HPV and LD on facial expressions recognition. In section ??, representative frameworks of automatic facial expression analysis systems are introduced. Representative databases used for facial expressions recognition are reviewed in section ?. A focus on these two challenges and the performances of several approaches are compared. A common experimental framework using a newly created data collection covering simultaneously free (camera in front of the subject) and constrained (camera attached to the head) facial expressions is proposed in section ?. A series of experiments are presented in section ??, in order to quantify the performance degradation induced by HPV and LD considering representative state-of-the-art approaches. In section ??, we summarize the limits of existing methods and data collections, as well as the benefits brought by the proposed experimental framework.

2. Large displacements (LD) and head pose variations (HPV)

In interaction situations, facial expression analysis has to deal with HPV and LD challenges. LDs involve translation, cinematic blur and scale changes, whereas, HPVs involve 3D-rotations (in-plane and out-of-plane). A first encountered issue with HPV is that most of the state-of-the-art approaches which give the best results in expression recognition are not invariant under 3D geometric transformations, thus computed features for the same face and the same expression vary depending on LD and HPV. For example, it is obvious that histogram-like [? ?] or dynamic

50 texture features computed from equal-sized facial grids are not invariant under translations, ro-
 51 tations and scale changes. Figure ?? shows an overview of a generic workflow often used in
 52 facial features extraction. Faces are divided into a regular grid of $m \times n$ local regions from which
 53 features can be extracted. Finally, features are concatenated into one-row vector which depicts
 54 the facial expression. HPV induces misalignment of the face (no correspondence of major facial
 55 components in each block, across the same facial image from a different point of view) and may
 56 results in mismatching between extracted features.

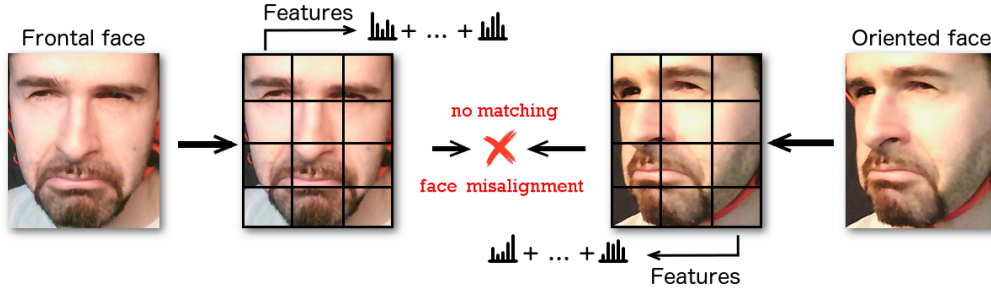


Figure 2: Example of misalignment of the face in the presence of head pose variations.

57 In order to obtain an invariance under geometric transformations, a pre-processing step which
 58 consists in registering faces is proposed in [? ?]. Face registration aims to find the transfor-
 59 mation (or the deformation) which reduces the discrepancies between two or more faces. These
 60 approaches modify facial characteristics (texture, geometry, motion) while reducing variations
 61 in translation, rotation and scale changes. However, registration induces artifacts which have a
 62 negative impact on the consistency of facial characteristics [?].

63 Another issue is encountered with LD which corresponds to important head motions between
 64 two frames. In the presence of LD, a blur effect appears on the face. This noise causes texture
 65 changes. Face registration suffers significantly under motion blur [?]. Indeed, most representa-
 66 tive face registration approaches are built on features (i.e facial landmarks), and their robustness
 67 is heavily dependent on the image quality and resolution. Hence, the performances of the regis-
 68 tration approaches may be less efficient when head motions occur. Figure ?? shows an example
 69 of mis-estimation of facial landmarks due to the blur effect caused by LD, which deteriorates the
 70 face registration.

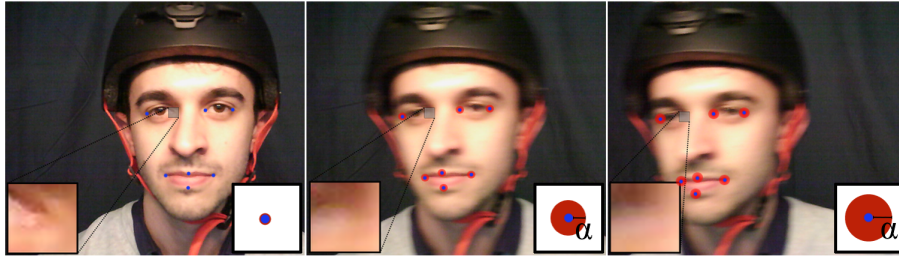


Figure 3: Poor estimation of landmarks location due to the blur effect caused by LD. α corresponds to the level of uncertainty concerning the landmarks location, which decreases sharply with the quality of the picture.

In brief, the presence of HPV and LD brings several challenges in the facial expression analysis :

- facial misalignment due to head pose variations
- preservation of initial facial expression during face registration process
- blur effect due to rapid movements resulting in poor landmark locations

In the next section, we discuss solutions to the challenges listed above.

3. Automatic facial expression analysis

Automatic facial expression analysis is a complex task as the face shape varies considerably from one individual to another. Furthermore, HPV and LD generate various face appearances for the same person. Such variations have to be addressed at different stages of an automatic facial expression analysis system. The generic facial expression analysis framework is illustrated in Figure ?? . First, the face is located in the frame and a registration step may be applied to remove the head motion and inter-subject differences. Next, the face is analyzed to estimate the remaining deformation caused by facial expressions. Then, features are extracted, and these features are used in the classification part of the system.

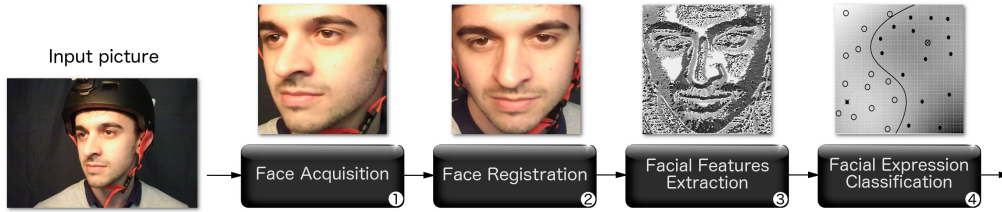


Figure 4: Generic facial expression analysis framework.

In the remainder of this section, we discuss the impact and the way HPV and LD are dealt with face registration and facial feature extraction processing stages.

3.1. Face registration approaches

The face is usually detected and registered in order to establish the correspondence of major facial components such as eyes, nose and mouth across different facial images. This aims at guaranteeing invariance to geometric transformations. In the following we discuss the benefits and limitations of various techniques such as eye-based registration (Eyes), as well as, more evolved techniques such as shape-based registration (Shape) or 3D model-based registration (3D Model).

Eyes registration. Eyes registration is the most popular strategy in near frontal-view databases [? ? ?]. Eyes are detected and images are aligned and scaled with regard to the inter-pupilar distance and orientation. Eyes are the most reliable facial component to be detected and suffer little changes in the presence of expressions. The limitation of this approach is that eyes must be well-detected. Usually, when out-of-plane rotations appear, the eyes quickly disappear and additional deformations are induced, avoiding the detection of eyes.

101 *Shape registration.* Shape registration is based on 2D facial landmarks and aims at increasing ro-
 102 bustness to HPV. Extensions considering more landmarks is supposed to provide greater stability
 103 in case of individual poor landmark detections. Some approaches [? ?] only rely on landmark
 104 points located near the center of the face. The inner landmarks are mostly used to detect the
 105 face and estimate the head pose. However, these points are affected by facial deformations in the
 106 presence of facial expressions. Other approaches also take into account the contour of the face
 107 in order to exploit the information related to the geometry of the face [?]. The outer landmarks
 108 are less affected by facial deformations due to facial expressions, but they are difficult to locate
 109 in case of out-of-plane rotations. We can say that most of 2D-feature-based methods are suitable
 110 for the analysis of near frontal facial expressions in the presence of limited head motions. But,
 111 they do not cope well with difficulties brought to occlusions and out-of-plane rotation. Indeed,
 112 an image acquisition system provides only the projection of the observed scenes in a 2D plane.
 113 The projection only captures information available in front of the camera and loses out-of-plane
 114 information. Figure ?? illustrates a poor estimation of facial landmarks due to a yaw out-of-plane
 rotation, where the left part of the face disappears progressively as the face rotates.

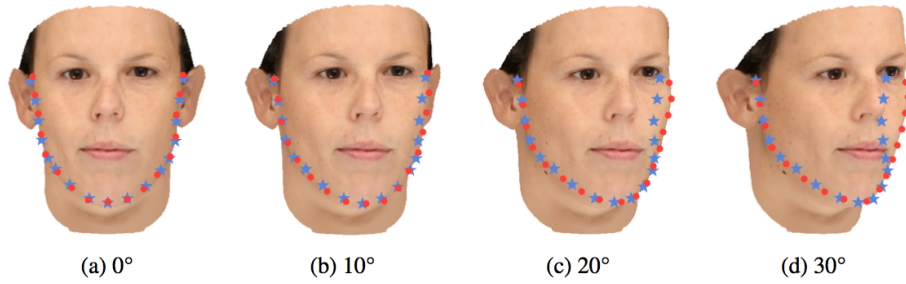


Figure 5: Similarity errors of 2D (red) and 3D (blue) facial contour landmarks under different angles [?].

115

116 *3D model registration.* Recent approaches propose robust face registration based on 3D to gen-
 117 erate a natural face image in frontal pose. Compared to 2D approaches, 3D approaches reduce
 118 the deformation of the face when facial expressions occur. Among these approaches, Zhu et al.
 119 [?] propose a robust face registration approach based on a 3D Morphable Model (3DMM). To
 120 build a 3D face model from a 2D face image, they estimate the depth of the external face region
 121 and the background.

122 **Pose registration** Landmarks are detected using facial alignment techniques from the 2D
 123 face. The authors apply landmark marching in order to solve the issue illustrated in Figure ??.
 124 Corrected landmarks on the boundary of the face are used as facial anchors. Facial anchors
 125 correspond to specific facial points that are used in order to align the 2D face on a 3D morphable
 126 model (constructed from large training data). A fitted 3D face is then generated and 3DMM
 127 coherently registers the face in front of the camera and preserves the appearance and the shape
 128 of the face. However, in case of high HPV, some regions can be hidden due to self-occlusions,
 129 as illustrated in Figure ??.

130 **Filling of occluded regions** Bad filling of the occluded region leads to large artifacts after
 131 registration and deteriorates recognition performance. To deal with self-occlusions, several ap-

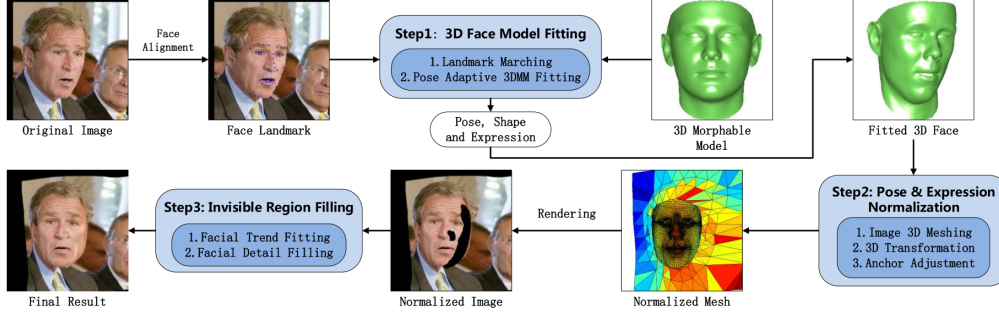


Figure 6: Face registration approach based on a 3D Morphable Model (3DMM), extracted from [?].

proaches use the facial symmetry or in-painting approaches [?]. The quality of these approaches depends on the size of the occluded face region and they are often not well-suited for use in unconstrained conditions (e.g illumination variation, occluding objects such as glasses and hair). Recent 3D approaches (such as [?]) use measurements over multiple frames to refine the rigid 3D shape and estimate hidden facial parts (assuming that the hidden facial part was visible in the previous frames).

In the next section, we present the most significant facial feature extraction approaches that have been proposed in the literature.

3.2. Facial Feature extraction approaches

In the literature, there are two types of methods used to analyze facial expressions : methods based on facial static characteristics and methods based on dynamic characteristics, each of them, applied locally or globally on the face.

Static texture features. Most of facial expression recognition approaches are based on the textural information [? ? ? ?]. One of the most popular method is the Local Binary Pattern (LBP) [?]. For every pixel of the image, its gray-scale value is compared with those of the eight surrounding pixels. The value of each neighbor is set to 0 if its gray-scale value is smaller than the value of the central pixel and to 1 otherwise. To reduce the dimensionality of the problem further, images are usually divided into a regular grid of $m \times n$ local regions from which LBP histograms can be extracted. Then, they are concatenated into a single histogram.

Dynamic features. The dynamic characterization of the facial texture can be achieved either by considering the changes in terms of temporal texture characteristics (extending static characteristics to temporal domain) or evaluating the changes in terms of perceived motion (by means of dense optical flow fields):

Region-based temporal texture features Dynamic texture is an extension of texture characterization to the temporal domain. Ambadar et al. [?] prove the importance of the dynamic texture for facial expression recognition as it allows a better analysis of physical deformation of face. Zhao et al. [?] propose an extension of the original LBP operator to the spatio-temporal domain called Volume Local Binary Patterns (VLBP). VLBP considers a block of video frames as a single 3 dimensional array of grayscale values. A simplified, more practical version of the

approach was proposed by its creators to make it more attractive for further usage called Local Binary Patterns from Three Orthogonal Planes (LBP-TOP) [?]. LBP-TOP applies LBP on every xy , x_t , and y_t slices separately. Then it averages the histograms over all slices in a single plane orientation, and concatenates the resulting histograms of the three dimensions. With LBP-TOP it is possible to combine motion and appearance analysis in one operator : the features histogram.

Optical flow features Optical flow measures the relative motion between two successive images in a sequence. It is used to analyze facial expression [? ?] and obtains good performances. Optical flow are dense and features encoding their local or global characteristics are extracted in order to exploit the encoded motion information. For instance, Histogram of Oriented Optical Flows (HOOF) feature [?] is successfully used in order to encode the distribution of optical flows and extract global movement characteristics. HOOF feature encodes the dense optical flow fields by cumulating directions binned with regard to the horizontal axis and by weighting their magnitude. The weighting step aims at minimizing the noise impact on the global feature. However, high HPV involves an important loss in terms of facial information and it reduces the recognition rate of facial expression algorithms. Indeed, occluded face areas of the current picture are defined by a set of pixels who disappear in the next picture when out-of-plane rotations occur. These pixels have no correspondence within the next picture. This results in motion that is not directly observable in these regions. Recent approaches use the boundaries of the face (which have a high probability of being occluded in case of HPV) in order to reduce the noise induced by motion discontinuities [?]. Therefore, they use fill-in methods based on the motion of the neighboring regions and the physical constraints of the face (wrinkles, shape, ...) [?].

Although dynamic textures approaches perform well in near frontal view, facial expression recognition based on dynamic textures, when HPV and LD occur, is still a challenging problem. Indeed, in these context stationary dynamic textures must be well-segmented in space and time. The performances of these approaches depend heavily on the quality of the face registration approaches to reduce facial deformations.

In the next section, we analyze how HPV and LD challenges are highlighted in several databases commonly used to validate facial expression analysis approaches.

4. Facial expression databases

Most of facial expression systems evaluate their performances in controlled settings [? ?], where the face pose is static. In these settings, expressions are exaggerated and often played by actors in order to induce important deformations on the face. In contrast, some data collections are recorded in more natural interaction contexts, where the subject has full freedom of its movement and facial expressions are spontaneous [? ? ?]. These databases, which propose more natural interactions, yield more often problems related to high HPV and LD. Some acted databases [? ?] have extended their data collections in order to offer a more challenging context for approaches aiming to improve their robustness to in-the-wild conditions.

The most commonly used databases for facial expressions analysis are shown in Figure ??.

Figure ?? shows that the complexity of the different databases increases depending on the type of expression (acted to spontaneous). Concerning the presence of LD and HPV, an indicator of intensity between one and three stars (✳) depicts the ratio of data which contains LD, or HPV.





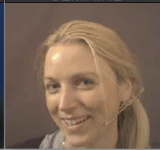
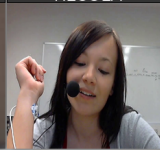
	CK+	MMI	GEMEP	DISFA	SEMAINE	RECOLA
						
Details	CK+ [28]	MMI [29]	GEMEP [4]	DISFA [31]	SEMAINE [5]	RECOLA [6]
Expressions	Acted	Acted	Acted	Spont.	Spont.	Spont.
Presence of HPV	-	*	**	*	**	***
Presence of LD	*	*	***	**	***	***

Figure 7: Commonly used databases for facial expression analysis.

Evolving challenges proposed in these databases reflect that the facial expression analysis in an interaction situation is a complex issue. In these contexts, the presence of HPV and LD challenges the current approaches.

Features previously discussed have been applied to several databases and the results are reported in Figure ?? . Results on CK+ [?] and MMI [?] show that facial expression analysis achieve excellent performances under controlled settings where the pose is static and expressions are acted. As illustrated in Figure ?? , registration approaches based on 3D models and 2D face shape are intensively used to analyze facial expressions in uncontrolled contexts. Despite the fact that these approaches provide better performances, recognition rates are still very low with regard to performances observed under controlled settings. However, in more natural interaction contexts like in GEMEP [?], DISFA [?] and SEMAINE [?], a significant drop in performance can be observed. To better visualize the performances of each registration approach on the various databases, we included in the lower part of Figure ?? two related graphics. A comparison of the approaches using average recognition or classification rates are given in Figure ??-A. Figure ??-B shows the performances of the approaches estimated using Person's cross correlation. Each color is associated with a pair of a registration approach and a database.

Recent databases, like SEMAINE [?] or RECOLA [?], include free head movements. Still, it is difficult to study the impact of head movements on facial expression recognition as many other parameters are changing within or between the existing databases. Hence, it is difficult to quantify the impact of issues related to LD and HPV, as well as, the registration techniques on the recognition performances. Basically, we are missing the equivalent near-frontal view data in order to measure effectively the induced deformations while correcting LD and HPV effects.

5. Synchronous acquisition system

In order to quantify the impact of free head movements on expression recognition performances, we propose an innovative acquisition system that collects data simultaneously in presence and absence of head movement. Experiments are then conducted in order to estimate the impact of HPV and LD on the recognition process.

To address this issue, we propose a new acquisition system called : Simultaneous Natural and Posed Facial expression (SNaP-2DFe) allowing the study of the HPV and LD impact on expression recognition methods.

References	Features	Registration	Databases		Performances
Zhao <i>et al.</i> [8]	LBP-TOP	Eyes	CK+	■	ar:95.2%
Happy <i>et al.</i> [2]	Salient LBP Patches	Eyes	CK+	■	ar:94.14%
Allaert <i>et al.</i> [25]	Optical flow & Geometry	Eyes	CK+	■	cr:95.34%
Koelstra <i>et al.</i> [3]	FFDs	Eyes	MMI	■	cr:94.3%
Jiang <i>et al.</i> [1]	LPQ-TOP	Shape	MMI	■	cr:94.7%
Rivera <i>et al.</i> [16]	DNG	Shape	MMI	■	ar:97.6%
Jiang <i>et al.</i> [22]	LPQ	N/A	GEMEP	■	cr:66%
Yang <i>et al.</i> [11]	LBP,LPQ	Shape	GEMEP	■	ar:84%
Sandbach <i>et al.</i> [13]	LBP	Eyes	DISFA	■	cc:0.342
Cruz <i>et al.</i> [32]	LPQ	Shape	SEMAINE	■	ar:55%
Nicolle <i>et al.</i> [33]	Appearance & Geometry	Shape	SEMAINE	■	cc:0.46
Chen <i>et al.</i> [34]	3D Facial Shape	3D Model	SEMAINE	■	cc:0.51

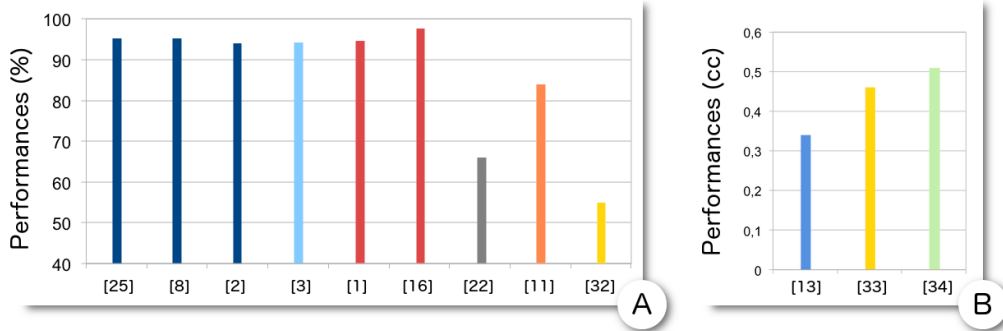


Figure 8: Recent methods to facial expression analysis in the literature. (cc: Person’s cross correlation, ar: average recognition rate, cr: classification rate).

5.1. Acquisition system

Each facial expression is recorded simultaneously using a two-camera system : one camera is fixed on a helmet, while the other is placed in front of the user at near-range distance. The helmet camera provides data similar to CK+ [?] and MMI [?] databases, where little or no head movements occur. The frontal camera provides data similar to RECOLA [?] and SEMAINE [?] databases, as subjects are freely moving their head. Our database enhances measuring the impact of head-movements relying on the information returned by the frontal camera, compared to the helmet camera.

The helmet is equipped with eight LEDs, which ensure homogeneous illumination on the face, even when the head is moving. It also includes a “9DOF Razor IMU” board by SparkFun, which contains a 3-axis gyroscope/accelerometer/magnetometer and a micro-controller performing sensor fusion. Finally, it includes a camera located fifty centimeters in front of the face and maintained by an aluminum rail in order to ensure global stability.

We use a counterweight that enhances the user’s comfort and guarantees that the helmet does not shift position while the user moves. It is important to guarantee that the helmet is stable in order not to disrupt the user experience during the recording session. We verify that the helmet is stable by computing the mean difference (in pixels) of facial landmark locations from the helmet

camera under neutral expression between different head poses. We have obtained very similar values regardless of the head movement. When the head is not moving we have obtained an error of 1.74 pixels. When the head is executing a diagonal movement we have obtained 1.87 pixels. In case of a Pitch, a Yaw or a Roll movement we have obtained respectively errors of 1.77 pixels, 1.77 pixels and 1.95 pixels. When the head is executing a translation an error of 1.71 pixels has been reported. With regard to the values, in our understanding, errors stem primarily from the instability of landmarks location detection and not the instability of the helmet.

The capturing system is illustrated in Figure ???. Each participant was instructed to wear a helmet fitted with a camera (Camera 1) and to sit in front of a projection screen at about one meter away from the fixed camera (Camera 2). We recorded images using two Creative Live cam inPerson HD (Full HD 1080p at a frame rate of 30 fps) and with an uniform background. The capturing system is illustrated in the left part of Figure ???. The right part of Figure ??? represents image samples of facial expressions where the subject performs a pitch movement. The first two lines correspond to selected synchronous frames in time. The first line corresponds to the helmet camera (Camera 1) and the second line, to the fixed camera (Camera 2). The curve in the bottom right part of the Figure ??? represents the yaw, pitch and roll values obtained by the gyroscope during the session.

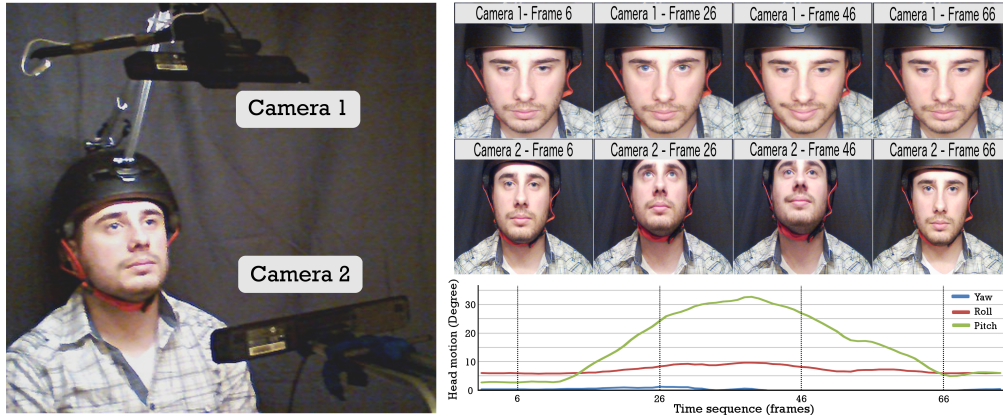


Figure 9: SNaP-2DFe system setup and example images of facial expressions recorded during a pitch movement.

5.2. Data acquired

Our preliminary database includes 840 samples collected from 10 subjects. Each video corresponds to a combination of one facial expression and a sequence of uniform head movements. In each sequence, the user follows a specific pattern of movement that corresponds to one of the following animations : one translation on x (T_x - corresponding to LD motions), combined with three rotations (roll, yaw, pitch - corresponding to HPV motions).

In static sequences, the user does not move the head, in order to collect data for head-posed facial expression analysis. In diagonal sequence, the user moves the head while combining translations and rotations. Movement ranges are large: translations go up to more than 150mm in any axis from the starting point, and rotations can reach 40 degrees.

For each subject, we have six animations combined with seven expressions (Neutral, Happiness, Sadness, Anger, Fear, Surprise, Disgust) from two cameras, which makes a total of 84 videos for each subject (42 with HPV and 42 without HPV).

The subjects were instructed to express emotions. Yet, as the subjects were not actors, in some situations, expressions recorded show spontaneous expressions characteristics: low intensity, limited facial deformations, various ways of expressing a given expression. In some other situations, the subjects were acting a different expression than the one they were asked for. However, this is not an issue at this point, as our main concern is to evaluate classification results comparing the near frontal-view data and the HPV-LD data of the same underlying expression.

The data collected is freely available for research purposes and can be downloaded on demand from <http://www.cristal.univ-lille.fr/FOX/>.

In order to assess the impact of face registration approaches on the recognition of facial expressions, the next section discusses the results of different registration methods on the collected SNaP-2DFe database.

6. Experimentation

Several experiments are conducted in the following. Firstly, we measure the ability of the registration techniques to simulate frontal pose images (like the ones produced by the helmet camera) from the static camera. As a reminder, the helmet-camera is a fixed frontal camera, where no head motion appears except the facial expressions. The characteristics of the face are stable during the sequence. This means that no registration step is necessary. In the experiments presented in Section ??, we evaluate the capabilities of face registration approaches to reduce the discrepancies between faces from frontal and non-frontal settings.

Secondly, we study the ability of registration techniques to preserve the original facial deformations produced by the underlying expression. In Section ??, expression recognition classifiers are trained from the helmet-camera images and we measure the ability of registration techniques to bring non-frontal images in frontal settings with regard to the frontal classifiers. In Section ??, we conduct a series of experiments, where the classifiers are trained from the registered images, in order to evaluate if the deformations, induced by the registration, preserve distinctive features for expression classification. Finally, we conduct an experiment in order to evaluate the impact of the registration deformation with regard to specific expressions in Section ??.

6.1. Evaluation of registration quality

In order to clearly illustrate the quality of the registration process we provide a qualitative and a quantitative evaluation. The qualitative evaluation illustrates visually the deformation induced by the registration process, whereas the quantitative evaluation measures the geometric and structural similarity between the registered face (from the fixed camera) and the near-frontal view face (from the helmet camera).

6.1.1. Qualitative evaluation

Figure ?? shows a qualitative comparison of three face registration techniques on different head poses extracted from SNaP-2DFe database (e.g. frontal pose, translation on x (Tx), roll, pitch, yaw and diagonal). To deal with near frontal face, Eyes registration is more adapted because this registration will not cause facial deformations and the locations of feature points are rather stable. However, severe out-of-plane rotation downgrade the precision of feature points

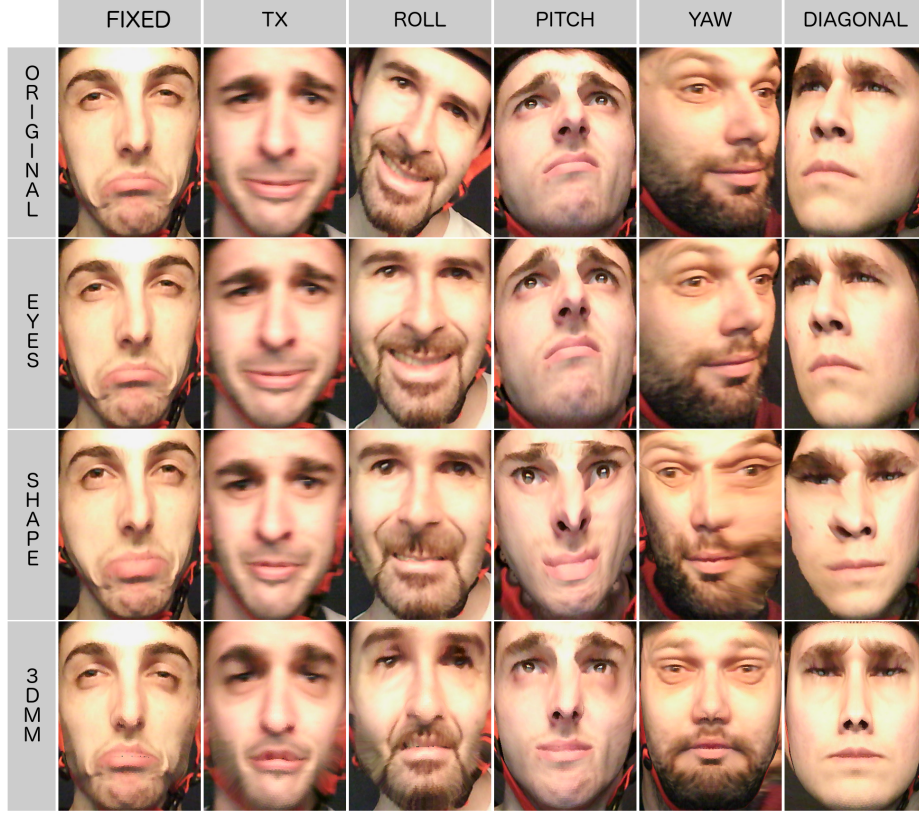


Figure 10: Example of different facial registration approaches on each animation.

319 localization process. This causes strong face deformations. In this case, recent approaches based
 320 on 3D face model seem better suited than other approaches. 3DMM method illustrated here is
 321 based on facial symmetry reconstruction [?]. Thanks to the reconstruction of occluded face
 322 regions, this approach allows rebuilding faces in the presence of out-of-plane rotations.

323 6.1.2. Quantitative evaluation

324 *Experimental Setup.* In order to evaluate the quality of face registration process, we use the
 325 structural similarity index method (SSIM) [?]. SSIM compares local patterns of pixel inten-
 326 sities that have been normalized for luminance and contrast. The face geometry delivers good
 327 information for some facial expressions, but fails in detecting subtle motions, that can be de-
 328 tected only by observing skin surface changes. From our point of view, SSIM is appropriate to
 329 measure the errors of the face registration for facial expression recognition. SSIM formula is
 330 based on three comparison measurements : luminance (l), contrast (c) and structure (s). The
 331 measure between two local regions x and y of common size $N \times N$ is :

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot s(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_x\sigma_y + c_2)(2cov_{xy} + c_3)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)(\sigma_x + \sigma_y + c_3)}. \quad (1)$$

where μ_x and μ_y are the average of x and y , σ_x^2 and σ_y^2 are the variance of x and y , and cov_{xy} is the covariance of x and y . Three variables $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$, $c_3 = c_2/2$ stabilize the division with weak denominator, where L corresponds to the dynamic range of the pixel-values and $k_1 = 0.01$ and $k_2 = 0.03$.

SSIM is applied on each animation of SNaP-2DFe, after using different registration approaches, based on : Affine Transformation on eyes location (Eyes), facial shape deformation using moving least squares [?] (Shape) and 3D Morphable Model [?] (3DMM). Results are reported in Table ??.

Result Analysis. 3DMM registration approach gives the best overall results, with mean similarity of 61.33% over all animations. Eyes and Shape registrations are very similar. Eyes registration approaches suit better in-plane geometric transformation (fixed, translation and roll) than Shape registration. Both present limitations due to the fact that they only exploit the visible 2D information, whereas 3DMM registration achieves better results in out-of-plane conditions (pitch, yaw, diagonal). In all cases, registration approaches improve the SSIM metrics in challenging conditions even though it does not guarantee a perfect match between the two cameras.

Registration	Fixed	Tx	Roll	Pitch	Yaw	Diagonal	Mean
None	53.30	47.88	48.68	46.34	46.54	51.14	48.98
Eyes	58.26	55.06	55.29	54.57	52.01	55.79	55.16
Shape	55.25	53.49	52.29	57.40	54.44	57.67	55.09
3DMM	64.72	61.81	58.17	60.52	58.47	64.29	61.33

Table 1: SSIM (in percentage) applied on each animation, with different face registration approaches.

Considering the results, face registration approaches may not ensure a perfect similarity between faces. But still, we expect that they encode expression-related artifacts that might still differentiate between expressions.

In the next section, we study the impact of the facial deformation induced by face registration approaches to facial expression recognition.

6.2. Evaluation of registration impact on expression recognition

We provide experimental results about the impact of facial registration on expression recognition performance when free head movements occur.

In this context, we try to measure the capacity of the registration method to induce facial deformation that can cope with classifiers learnt from the near-frontal recordings provided by the helmet camera.

Experimental Setup. The next series of experiments are conducted using LIBSVM [?] with the Radial Basis Function kernel for classification. Each expression is classified into one of the seven classes : Neutral, Anger, Fear, Disgust, Happiness, Sadness, and Surprise. Assuming the face region is well aligned after applying different face registration approaches, we use a 5*5 facial block based approach to extract the features. We consider commonly used static (LBP) and dynamic (LBP-TOP, HOOF) texture features for classifying expressions. The implementation of LBP and LBP-TOP were taken from [?]. HOOF feature extraction was reimplemented by us considering the algorithm described in [?]. Expression recognition rates are computed by employing the 10-fold cross-validation protocol.

In the first experiment, near-frontal faces recorded with the helmet camera are used for the training step. This first experiment allows to evaluate the performance of the different approaches in good conditions (not involving HPV or LD). The characteristics of the face are stable during the sequence, hence no registration step is necessary. All registration approaches were applied on each animation. The expression recognition classifier is trained using images captured in near-frontal settings using the helmet camera.

Result Analysis. The results for the different configurations : fixed versus helmet camera, no registration versus various registration approaches on the fixed camera are given in Table ?? . A first look at the results obtained in Table ?? shows that the originating camera and face registration approaches have significant impact on the performances. In the following we discuss the impact of the features, the registration method as well as the originating camera.

	Helmet camera	Fixed camera			
Method	Original data	Original data	Eyes	Shape [?]	3DMM [?]
LBP	75.52	30.55	47.46	47.76	51.34
LBP-TOP	78.34	19.44	49.12	44.62	46.93
HOOF	83.21	17.38	50.01	42.16	48.73

Table 2: Facial expression recognition rates while using different face registration approaches.

Impact of the originating camera Results reported in first column of Table ?? show that the state-of-the-art methods are suitable for the analysis of facial expressions when the head is not moving. However, in the presence of HPV and LD, images provided by the fixed camera are poorly classified as shown in the second column (Fixed Camera - Original Data).

Impact of the features The results obtained with the helmet camera show that dynamic texture features such as LBP-TOP or HOOF are more efficient than LBP. The HOOF approach obtains better performances than LBP-TOP where little or no head movement appears and proves that optical flow approaches are better suited to the facial expression analysis. However, the experiment shows a drastic fall in performances on the original data from the fixed camera. In this context, recognition rates measured with dynamic texture features have suffered more than others. Overall, these methods are much more sensitive to the presence of HPV and LD. It is important, therefore, to ensure that the face is aligned in order to maintain the benefits brought by the dynamic texture features.

Impact of the registration method When considering results obtained using various registration methods on the images captured with the fixed camera (last three columns), it can be easily seen that the performances are very similar. Each column corresponds to the expression classification rate obtained after applying a different registration approach, considering respectively : affine transformation on eyes location (Eyes), facial shape deformation using moving least squares [?] (Shape) and 3D Morphable Model [?] (3DMM).

The use of registration techniques improves significantly the performances of facial expression analysis when free head movements occur. Considering the results in Table ?? , the Eyes registration seems to be the most successful strategy in terms of sustainability and effectiveness

with regard to the needs of facial expression classification method. Despite the gain obtained from both Shape and 3DMM registration approaches when using LBP, these registration techniques appear to be less suited with dynamic texture features.

6.3. Evaluation of preserving distinctive facial expression deformations

In the following, we evaluate the impact of specific head motion patterns on the expression recognition rates when using LBP features. Implicitly, we quantify the capacity of registration techniques to preserve distinctive facial deformations in case of various movements typologies.

Experimental setup. The selected registration techniques have been applied and compared on each class of animations. We evaluate the impact of face registration approaches on facial expressions recognition rates and we identify the strengths and weaknesses of each. The results are given in Table ???. The training was performed on the registered images captured by the fixed camera. Hence, the trained classifier took into account the deformation induced by the registration.

Registration	Fixed	Tx	Roll	Pitch	Yaw	Diagonal
Original data	45.23	38.09	32.47	37.71	33.33	14.26
Eyes	52.38	33.33	47.61	30.95	40.47	26.19
Shape	47.61	35.71	42.85	30.95	38.02	11.90
3DMM	48.02	39.27	40.21	40.74	41.08	34.96

Table 3: Recognition rates of facial expression classification using LBP features, after face registration step.

Results analysis. The results in bold in Table ??? show the best results per registration approach obtained for specific movement patterns. While the head does not perform out-of-plane rotations, as in Fixed and Roll settings, Eyes registration provides the best results. Whereas, in case of out-of-plane rotations the 3DMM registration performs better.

The Eyes registration is the most suitable in frontal (fixed) settings. Indeed, in near-frontal view condition, a simple in-plane rotation aligns the face. This solution preserves the geometry of the face. However, this method is not working well in case of LD or HPV. Thanks to the reconstruction of the occluded face region, 3DMM approach obtains the best results when out-of-plane rotations occur. However, the reconstruction system is based on a face mirroring technique, which sometimes has negative impact on the induced facial expression.

6.4. Evaluation of per-expression registration impact

We have conducted a complementary study about measuring the impact of the registration techniques with regard to the underlying expression.

Experimental setup. In the light of the previous results (see Table ???), we have selected 3DMM and Shape registrations, as well as, LBP as texture features for studying the recognition rate variations when considering various expression classes. We have constructed ROC curves using 10-fold cross-validation protocol considering the whole dataset, as well as, independent Neutral, Anger, Surprise, Happiness, Sadness, Fear and Disgust partitions. Training was conducted on the whole dataset, as well as on each expression-related partition resulting in eight different classifiers.

433 *Results analysis.* Figure ?? shows ROC curves corresponding to each expression score calculated with LBP from the helmet camera (blue) and the fixed camera after different registrations (red : 3DMM and green : Shape). With regard to the Mean Curve, faces obtained by registration show lower performances than the faces acquired by the helmet camera (see Mean Curve in Figure ??).

438 Results show that some expressions suffered severely from the registration process. Expressions like Anger, Surprise are less impacted by facial registration. This is probably due to the fact that face registration process induces less facial deformations around regions (such as eyebrows) used in the recognition process. Disgust and Fear expressions show similar behavior as Anger and Surprise, but the 3DMM registration technique seems more robust. Expressions like Happiness, Neutral and Sadness seem more impacted by the registration as regions outside the landmarks are affected (such as upper cheeks for Happiness). The drop in performances in the case of Neutral expressions underlines the fact that the deformations produced by the registration induce "false" facial expression recognition.

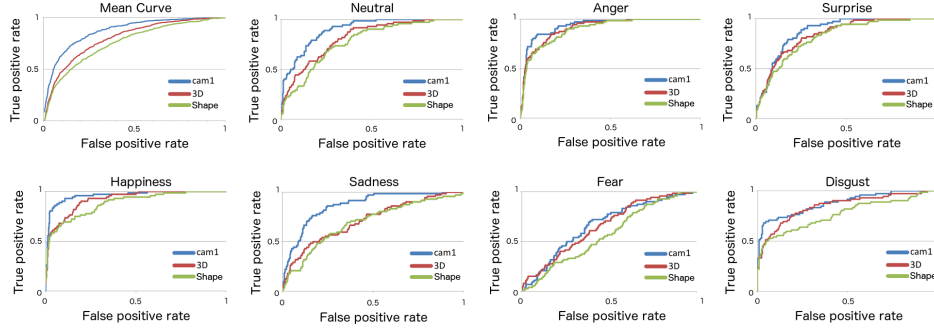


Figure 11: ROC curves for the 10-fold cross-validation protocol for LBP.

447 In the light of results from these experiments, the facial registration approaches improve the facial expression recognition in case of HPV and LD. Yet, a lot of improvements still have to be made in order to obtain comparable performances in the two settings : near-frontal views vs unconstrained head movements. The choice of face registration approach is heavily dependent on the type of head motion variation occurring in the video sequence. Mainly used for face recognition, these registration approaches do not ensure a convenient alignment persevering the expression of the face over time.

454 In the following section, we summarize the contributions of this study. From literature results (see section ??), and results obtained in our experimental settings we discuss perspectives concerning LD and HPV.

457 7. Conclusion and future works

458 In this paper we have addressed the study of the impact of registration techniques on expression recognition performances. Registration techniques are employed in order to handle HPV and LD for facial expression recognition. When analyzing the facial characteristics (texture, geometry, motion) for expression recognition, facial distortions due to misalignment degrade the

performances of the system. Removing distortions is a complex task. Most of the time it has a negative impact on the coherency of facial characteristics (texture, geometry, motion) [?].

3D Model registration is constantly improving with regard to in-the-wild challenges, but there is still no solution to ensure a satisfactory face registration while maintaining the facial expression. Indeed, the use of face registration techniques does not seem adequate to preserve the features encoding facial expression deformations. The loss in terms of precision when considering free head movements is partly due to the noise induced by the face registration process itself.

In this paper, we propose an innovative acquisition system, in order to quantify the impact of free head movements on expression recognition performances. Experiments on the impact of well known head pose registration techniques (Eyes, Shape or 3DMM) on facial expression recognition are reported. The results show that the face registration and the facial expression recognition approaches are heavily dependent on the type of head motion variation. When considering static approaches (such as LBP), in the presence of in-plane rotations, registration techniques based on landmarks (such as Eyes or Shape) preserves better the underlying expression. However, when out-of-plane rotations occur, registration techniques based on the reconstruction of 3D models seem more accurate as they preserve the underlying expressions better. Approaches using dynamic features (such as LBP-TOP, HOOF) are more efficient in terms of facial expression analysis for frontal poses. However, these approaches do not handle well face registration techniques (Eyes, Shape or 3DMM).

Out-of-plane rotations affect in a strong manner, the expressions recognition process. Supporting out-of-plane rotations can be achieved either by incrementing data (as in Deep-learning methods) or by registering the face representation to near frontal views. Although the first approach seems more popular at the present time, we truly believe that progress can be made in the latter by creating innovative face registration techniques that preserve facial expression. The SNaP-2DFe database can jointly be used to propose and evaluate innovative registration techniques while reinforcing the facial expression recognition or the head pose estimation methods.

As an alternative to registration techniques, we think that solutions from the field of dense optical flow should be explored. The enhancements of post-filtering solutions capable of registering the movement, by filtering out the head movement and keeping only the inner-facial local movement could be done. The database proposed here may serve future works in this direction. The helmet camera provides the ground truth movements while the static camera provides the challenging data from where the head movements should be subtracted by future post-filtering solutions.

8. References